



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

A skew Gaussian decomposable graphical model

Citation for published version:

Zareifard, H, Rue, H, Khaledi, MJ & Lindgren, F 2016, 'A skew Gaussian decomposable graphical model' Journal of Multivariate Analysis, vol 145, pp. 58-72. DOI: 10.1016/j.jmva.2015.08.011

Digital Object Identifier (DOI):

[10.1016/j.jmva.2015.08.011](https://doi.org/10.1016/j.jmva.2015.08.011)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Journal of Multivariate Analysis

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



A skew Gaussian decomposable graphical model

Hamid Zareifard^a, Håvard Rue^b, Majid Jafari Khaledi^c, Finn Lindgren^d

^a Department of Statistics, Jahrom University, Jahrom, Iran

zareifard@jahrom.ac.ir

^b Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway

hrue@math.ntnu.no

^c Department of Statistics, Tarbiat Modares University, P.O. Box 14115-134, Tehran, Iran

jafari-m@modares.ac.ir

^d Department of Mathematical Sciences, University of Bath, Claverton Down, Bath, BA2 7AY, United Kingdom

f.lindgren@bath.ac.uk

Abstract

This paper propose a novel decomposable graphical model to accommodate skew Gaussian graphical models. We encode conditional independence structure among the components of the multivariate closed skew normal random vector by means of a decomposable graph and so that the pattern of zero off-diagonal elements in the precision matrix corresponds to the missing edges of the given graph. We present conditions that guarantee the propriety of the posterior distributions under the standard noninformative priors for mean vector and precision matrix, and a proper prior for skewness parameter. The identifiability of the parameters is investigated by a simulation study. Finally, we apply our methodology to two data sets.

Keywords: Decomposable graphical models; multivariate closed skew normal distribution; Conditional independence; Noninformative prior.

1 Introduction

In recent years, there have been many developments in multivariate statistical models. Making sense of all the many complex relationships and multivariate dependencies present in the data, formulating correct models and developing inferential procedures is an important challenge in modern statistics. In this context, graphical models currently represent an active area of statistical research which have served as tools to discover structure in data. More specifically, graphical models are multivariate statistical models in which the corresponding joint distribution of a family of random variables is restricted by a set of conditional independence assumptions,

and the conditional relationships between random variables are encoded by means of a graph. In the Gaussian case, these models induce the conditional independence assumptions by zeros in the precision matrix. An important reason for working with this class of distributions is important properties like closure under marginalization, conditioning and linear combinations which is seldom preserved outside the class of multivariate normal distributions. However, in spite of substantial advances, the Gaussian distributional assumption might be overly restrictive to represent the data. The real data could be highly non-Gaussian and may show features like skewness.

In this article, we study skew distributions in graphical models with the aim of mimicking the success of Gaussian graphical models as much as possible. The last decade has witnessed major developments in models whose finite dimensional marginal distributions are multivariate skew-normal. Azzalini and Capitanio (1999) introduced multivariate skew-normal (SN) distribution which enjoys some of the useful properties of normal distribution, such as property of closure under marginalization and conditioning. Accordingly, an n -dimensional random vector \mathbf{Y} is said to have a SN distribution if its density is

$$\phi_n(\mathbf{y}; \boldsymbol{\mu}, \Omega) \Phi(\alpha_0 + \boldsymbol{\alpha}' D_{\Omega}^{-1}(\mathbf{y} - \boldsymbol{\mu})) / \Phi(\tau),$$

where $\phi_n(\cdot; \boldsymbol{\mu}, \Omega)$ is the probability density function of the n -dimensional $N_n(\boldsymbol{\mu}, \Omega)$ variable, $\Phi(\cdot)$ is cumulative distribution function of $N(0, 1)$, $\boldsymbol{\mu} \in \mathbb{R}^n$, $\tau \in \mathbb{R}$, $\Omega \in \mathbb{R}^{n \times n}$ is a full rank covariance matrix, $D_{\Omega} = \text{diag}(\Omega_{11}, \dots, \Omega_{nn})^{1/2}$, $\boldsymbol{\alpha} \in \mathbb{R}^n$ is shape parameter and $\alpha_0 = \tau(1 + \boldsymbol{\alpha}' D_{\Omega}^{-1} \Omega D_{\Omega}^{-1} \boldsymbol{\alpha})^{1/2}$. When $\boldsymbol{\alpha} = \mathbf{0}$ we are back to the multivariate normal distribution. Capitanio et al. (2003) used the SN family in graphical models examining in particular the construction of conditional independence graphs. Their results show that if \mathbf{Y} be an n -variate SN distribution with covariance matrix Ω and skewness vector $\boldsymbol{\alpha}$, then

$$Y_i \perp Y_j | \mathbf{Y}_{-ij} \Leftrightarrow \Omega^{ij} = 0 \text{ and } \alpha_i \alpha_j = 0$$

where \mathbf{Y}_{-ij} is \mathbf{Y} with the i th and j th elements deleted and Ω^{ij} denotes the (i, j) th entry of Ω^{-1} . Comparing with the Gaussian graphical model, an extra constraint $\alpha_i \alpha_j = 0$ is necessary to capture conditional independence property. It means if we believe Y_i and Y_j are conditionally independent then at least one of α_i and α_j must be zero. Hence, applying this constraint in practical issues is challenging. Alternatively, Dominguez-Molina et al. (2003) and Gonzalez-Farias et al. (2004) proposed the multivariate closed skew normal (CSN) distribution which includes the property of SN family. Also unlike the SN family, the CSN family enjoys this property that the joint distribution of i.i.d. CSN random variables is the multivariate CSN distribution.

Although much progress has been made in the context of skew normal distributions, the achieved successes in graphical models are limited. The preservation of conditional independence property for skew normal variables has shown that the extending the class of skew-normal distributions to graphical models is challenging. The aim of this paper is to develop a multivariate closed skew normal graphical model. We encode conditional independence structure among the components of the multivariate closed skew normal random vector with respect to a decomposable graph G . The main motivation to use the decomposable graphs for encoding

the conditional independence is that for this type of graph there exists an ordering of the vertices such that the zero elements in precision matrix are reflected in its Cholesky decomposition (Paulsen et al., 1989). Under decomposable graphs, the conditional independence property is maintained for our skewed graphical model, and simplification occurs in both the interpretation of data and the estimation procedure. Models can be specified in terms of conditional and marginal probability distributions, leading to a simplified analysis based on lower dimensional components (Giudici and Green, 1999; Letac and Massam, 2007; Khare and Rajaratnam, 2011).

The precision matrix is the fundamental object that evaluates conditional dependence between random variables. Estimating a sparse precision matrix is crucial specially in high-dimensional problems. In this context, a family of conjugate prior distributions for the precision matrix is developed. Conditions for propriety of the posterior are given under the standard noninformative priors on mean vector and precision matrix as well as a proper prior for the skewness parameter. We also develop and implement a Markov chain Monte Carlo (MCMC) sampling approach for inference.

The organization of the paper is as follows. Section 2 introduces the required preliminaries and notation. In Section 3, a novel skew Gaussian decomposable graphical model is constructed using a multivariate closed skew normal distribution and its properties are established. Section 4 discusses Bayesian analysis using Gibbs sampling to sample from the posterior distribution. A simulation study is reported in Section 5. Section 6 illustrates the use of proposed methodology in two real data sets: an analysis of student marks from Mardia et al. (1979) and an analysis of the carcass data from gRbase package of R. Finally, conclusions and discussion are given in Section 7. The Appendix contains proofs of some of the results in the main text.

2 Preliminaries

2.1 Multivariate closed skew-normal distribution

An n -dimensional random vector \mathbf{Y} is said to have a multivariate closed skew-normal distribution, denoted by $CSN_{n,m}(\boldsymbol{\mu}, \Sigma, \Gamma, \boldsymbol{\nu}, \Delta)$, if its density function is of the form

$$f(\mathbf{y}) = \phi_n(\mathbf{y}; \boldsymbol{\mu}, \Sigma) \Phi_m(\Gamma(\mathbf{y} - \boldsymbol{\mu}); \boldsymbol{\nu}, \Delta) / \Phi_m(0; \boldsymbol{\nu}, \Delta + \Gamma \Sigma \Gamma'), \quad (1)$$

where $\boldsymbol{\mu} \in \mathbb{R}^n$, $\boldsymbol{\nu} \in \mathbb{R}^m$, and $\Sigma \in \mathbb{R}^{n \times n}$ and $\Delta \in \mathbb{R}^{m \times m}$ are both covariance matrices, $\Gamma \in \mathbb{R}^{m \times n}$, and $\Phi_n(\cdot; \boldsymbol{\mu}, \Sigma)$ is the cumulative distribution function of the n -dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ . To derive this distribution, Gonzalez-Farias et al. (2004) consider a $(n+m)$ -dimensional normal random vector

$$(W_{01}, \dots, W_{0m}, W_1, \dots, W_n)' = \begin{pmatrix} \mathbf{W}_0 \\ \mathbf{W} \end{pmatrix} \sim N_{n+m} \left(\mathbf{0}, \begin{pmatrix} \Delta + \Gamma \Sigma \Gamma' & \Gamma \Sigma \\ \Sigma \Gamma' & \Sigma \end{pmatrix} \right).$$

Then, the probability density function of $\mathbf{Y} = (\mathbf{W} | \mathbf{W}_0 > \boldsymbol{\nu})$ is the multivariate closed skew-normal (1). If $\Gamma = 0$ and $m = 1$, this density reduces to the multivariate

normal one and the skew-normal distribution (Azzalini, 2005), respectively. Allard and Naveau (2007) used the multivariate closed skew normal and introduced a spatial skewed Gaussian process as a novel way of modeling skewness for spatial data. To increase the amount of skewness in the vector \mathbf{Y} as well as to simplify the interpretation of this density, they assumed that $m = n$, $\boldsymbol{\nu} = 0$, $\Delta = \Sigma$ and $\Gamma = \alpha I_n$ in which $\alpha \in \mathbb{R}$ is a single parameter controlling skewness and I_n is the identity matrix of order n . This model is also referred to as the homotopic model.

We will make use of the alternative representation of Dominguez-Molina et al. (2003) for CSN families,

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + (\Sigma^{-1} + \Gamma' \Delta^{-1} \Gamma)^{-\frac{1}{2}} \mathbf{V} + \Sigma \Gamma' (\Delta + \Gamma \Sigma \Gamma')^{-1} \mathbf{U}, \quad (2)$$

where \mathbf{V} and \mathbf{U} are independent multivariate normal distribution $N_n(0, I_n)$ and a multivariate truncated normal distribution $TN_n(0; 0, \Delta + \Gamma \Sigma \Gamma')$, respectively, where $TN_n(\mathbf{c}; \boldsymbol{\mu}, \Sigma)$ denotes the $N_n(\boldsymbol{\mu}, \Sigma)$ distribution truncated below at the vector \mathbf{c} . Also, if \mathbf{Y} is partitioned as $\mathbf{Y} = (\mathbf{Y}'_1, \mathbf{Y}'_2)'$, then the conditional distribution of $(n - k)$ -dimensional vector \mathbf{Y}_2 given $\mathbf{Y}_1 = \mathbf{y}_{1_0}$ is

$$CSN_{n-k,m}(\boldsymbol{\mu}_2 + \Sigma_{21} \Sigma_{11}^{-1} (\mathbf{y}_{1_0} - \boldsymbol{\mu}_1), \Sigma_{22.1}, \Gamma_2, \boldsymbol{\nu} - \Gamma^* (\mathbf{y}_{1_0} - \boldsymbol{\mu}_1), \Delta),$$

where $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$, $\Gamma^* = \Gamma_1 + \Gamma_2 \Sigma_{21} \Sigma_{11}^{-1}$ and the parameters are induced corresponding to the partition of \mathbf{Y} as

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad \Gamma = \begin{pmatrix} \Gamma_1 & \Gamma_2 \end{pmatrix}.$$

2.2 Graph theory

Let V be a finite set of vertices and $E = \{(u, v) : u, v \in V, u \neq v\}$ be a set of edges so that $E \subseteq V \times V$. Define a graph G as an ordered pair $G = (V, E)$ of vertices and edges, where V is assumed to be finite. When $(u, v) \in E$, we say that u and v are adjacent in G . A graph is said to be complete if all the vertices are adjacent to each other. It is understood that $(u, v) \in E$ implies $(v, u) \in E$, i.e., the edges are undirected. For an undirected graph $G = (V, E)$, we say that u and v are neighbors when $(u, v) \in E$. For any $A \subset V$, a subgraph of V is defined as the graph $G_A = (A, E \cap (A \times A))$. A path from v_1 to v_m is a sequence of distinct vertices in V , v_1, v_2, \dots, v_m , for which $(v_j, v_{j+1}) \in E$ for $j = 1, \dots, m - 1$. A subset $C \subset V$ separates two vertices $i \notin C$ and $j \notin C$, if every path from i to j contains at least one vertex from C . A clique of G is a complete subgraph of G . A subgraph is a maximal clique if it is not contained in a larger complete subgraph. Any path that begins and ends at the same vertex is called a cycle. A tree is a connected graph with no cycles.

Definition 1: (Decomposable graph (Lauritzen, 1996)) An undirected graph is said to be decomposable if any induced subgraph does not contain a cycle of length greater than or equal to four.

Figure 1 shows examples of non-decomposable and decomposable graphs. Decomposable graphs have several characterizations in terms of vertex orderings.

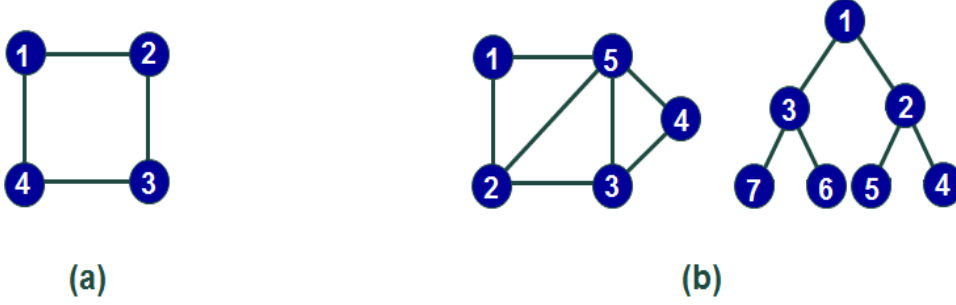


Figure 1: (a) A non-decomposable graph, and (b) two decomposable graphs.

Definition 2: For an undirected graph $G = (V, E)$, an ordering $\gamma : V \mapsto \{v_1, \dots, v_m\}$ of m vertices $(1, \dots, m)$ is known as a perfect vertex elimination scheme for G if for every triplet v_i, v_j, v_k with $1 \leq i < j < k \leq m$ the following condition holds:

$$(\gamma^{-1}(v_j), \gamma^{-1}(v_i)) \in E, \quad (\gamma^{-1}(v_k), \gamma^{-1}(v_i)) \in E \quad \Rightarrow \quad (\gamma^{-1}(v_k), \gamma^{-1}(v_j)) \in E.$$

In Figure 1(b), we show examples of perfect vertex elimination schemes for some decomposable graphs. The existence of such an ordering is an important advantage of decomposable over nondecomposable graphs, and the existence of this ordering characterize decomposable graphs. More formally, every decomposable graph admits an ordering of vertexes in terms of its cliques which is a perfect vertex elimination scheme. If for an undirected graph G , there exists an ordering of its vertices corresponding to a perfect vertex elimination scheme, then G is a decomposable graph (Lauritzen, 1996, page 18). However, this ordering need not be unique. A constructive way to obtain such an ordering is given in Lauritzen (1996).

Definition 3: (Modified Cholesky decomposition) If Q is a positive definite matrix, then there exists a unique decomposition

$$Q = L'DL$$

where L is a upper triangular matrix with unit diagonal entries and D a diagonal matrix with positive diagonal entries. The Cholesky decomposition of Q is $\Gamma'\Gamma$ where $\Gamma = D^{\frac{1}{2}}L$ is called the Cholesky triangle.

The following lemma will play a central role in our work.

Lemma 1: (Paulsen et al. 1989) Let Ω be an arbitrary positive definite matrix with zero restrictions according to decomposable graph $G = (V, E)$, i.e., $\Omega_{ij} = 0$ whenever $(i, j) \notin E$. Then there exists an ordering of the vertices such that if $\Omega = L'DL$ is the modified Cholesky decomposition corresponding to this ordering, then for $i < j$,

$$L_{ij} = 0 \Leftrightarrow (i, j) \notin E.$$

Hence, the zeros in Ω are preserved in the lower triangle of the corresponding matrix L obtained from the modified Cholesky decomposition. We assume from now on that graph G is decomposable.

An m -dimensional Gaussian graphical model can be represented by the class of multivariate normal distributions with fixed zeros in the precision matrix (i.e., conditional independencies) described by a given graph $G = (V, E)$ where the number of vertices is m . That is, if $(i, j) \notin E$, the i th and j th components of the multivariate random vector are conditionally independent.

We now introduce some notations and spaces from Khare and Rajaratnam (2011). Let M_m^+ and M_m denote the cone of symmetric positive definite matrices of order m and symmetric positive semidefinite matrices of order m , respectively, then the two parameter sets P_G^m and SP_G^m according to decomposable graph $G = (V, E)$ are defined as

$$\begin{aligned} P_G^m &= \{\Omega \in M_m^+ \mid \Omega_{ij} = 0, \ (i, j) \notin E\}, \\ SP_G^m &= \{\Omega \in M_m \mid \Omega_{ij} = 0, \ (i, j) \notin E\}, \end{aligned}$$

and also define

$$\begin{aligned} \ell_G^m &= \{L : L_{ij} = 0 \text{ whenever } i > j, \text{ or } (i, j) \notin E \text{ and } L_{ii} = 1 \ \forall \ 1 \leq i, j \leq m\}, \\ \Theta_G^m &= \{\theta = (L, D) : L \in \ell_G^m \text{ and } D \text{ diagonal with } D_{ii} > 0 \ \forall \ 1 \leq i \leq m\}. \end{aligned}$$

3 Skew Gaussian decomposable graphical models

To introduce a skew version of k -dimensional Gaussian graphical (GG) model, we first consider the following tree graphical model in a loose and imprecise form

$$\overset{1}{\bigcirc} \text{---} \overset{2}{\bigcirc} \text{---} \overset{3}{\bigcirc}. \quad (3)$$

Based on this graph, let $\mathbf{X} = (X_1, X_2, X_3)$ be a random vector whose elements are indexed by graph (3). We introduce three independent increments

$$\epsilon_1 = \kappa_1 X_1, \quad \epsilon_2 = \kappa_2 (X_2 - b_{12} X_1), \quad \epsilon_3 = \kappa_3 (X_3 - b_{23} X_2).$$

where $\kappa_1, \kappa_2, \kappa_3 \in \mathbb{R}^+$ and $b_{12}, b_{23} \in \mathbb{R}$. Now, we can relate $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \epsilon_3)$ to $\mathbf{X} = (X_1, X_2, X_3)$ by $\boldsymbol{\epsilon} = D_{\kappa}^{\frac{1}{2}} B \mathbf{X}$ where

$$B = \begin{pmatrix} 1 & 0 & 0 \\ -b_{12} & 1 & 0 \\ 0 & -b_{23} & 1 \end{pmatrix}, \quad \text{and} \quad D_{\kappa} = \text{diag}(\kappa_1^2, \kappa_2^2, \kappa_3^2).$$

If we suppose $\epsilon_1, \epsilon_2, \epsilon_3 \stackrel{iid}{\sim} N(0, 1)$, then the joint density of $\mathbf{X} = (X_1, X_2, X_3)$ becomes trivariate normal distribution with fixed zeros in the precision matrix described by given graph. Our aim is to assess the suitability and wider applicability of asymmetric distributions in graphical models with the hope of mimicking the success of Gaussian graphical models as much as possible. To do this, we need to study the following two questions

- How can we choose a skewed distribution for $\boldsymbol{\epsilon}$ such that the distribution of \mathbf{X} belong to the same class as that of the $\boldsymbol{\epsilon}$?

- What conditions are necessary to achieve the Markov property for the skewed graphical model?

To address these questions, we will use the CSN model that is more general than SN, and is closed under marginalization, conditioning and linear combination. Also unlike the SN family the joint distribution of i.i.d. CSN random variables is a multivariate CSN distribution. We now assume

$$\epsilon_i \sim CSN_{1,1}(0, 1, \alpha_i, 0, 1), \quad i = 1, 2, 3 \quad \Leftrightarrow \quad \boldsymbol{\epsilon} \sim CSN_{3,3}(\mathbf{0}, I_3, D_\alpha, \mathbf{0}, I_3)$$

where $D_\alpha = \text{diag}(\alpha_1, \alpha_2, \alpha_3)$. Note that α_i is interpreted as skewness parameters of the i th increment. Using the closure property of the CSN distribution under linear transformations, the joint density for \mathbf{X} becomes

$$p(\mathbf{x}) = 2^3 \phi_3(\mathbf{x}; 0, (B' D_\kappa B)^{-1}) \Phi_3(D_\alpha B \mathbf{x}; \mathbf{0}, D_\kappa^{-1}).$$

In this example, we can easily show that $X_1 \perp X_3 | X_2$, so we have the Markov property. The question is then: how can we extend this easy example to a more general graphical model? This we will study next.

Definition 4: Consider an undirected decomposable graph $G = (V, E)$ where the number of vertices is k and ordering of the vertices corresponds to a perfect vertex elimination scheme. A random vector $\mathbf{X} = (X_1, \dots, X_k)'$ is called a skew Gaussian decomposable graphical (SGDG) model with respect to graph G with mean $\boldsymbol{\mu} \in \mathbb{R}^k$, the precision matrix $Q \in P_G^k$ and the skewness parameters $\boldsymbol{\alpha} \in \mathbb{R}^k$, if its density is $CSN_{k,k}(\boldsymbol{\mu}, Q^{-1}, D_\alpha L, \mathbf{0}, D_\kappa^{-1})$ or equivalently

$$\begin{aligned} p(\mathbf{x}) &= 2^k \phi_k(\mathbf{x}; \boldsymbol{\mu}, Q^{-1}) \Phi_k(D_\alpha L(\mathbf{x} - \boldsymbol{\mu}); \mathbf{0}, D_\kappa^{-1}) \\ &= \left(\frac{2}{\pi}\right)^{k/2} |D_\kappa|^{1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' Q (\mathbf{x} - \boldsymbol{\mu})\right) \Phi_k(D_\alpha L(\mathbf{x} - \boldsymbol{\mu}); \mathbf{0}, D_\kappa^{-1}), \quad (4) \end{aligned}$$

where $D_\kappa = \text{diag}(\kappa_1^2, \dots, \kappa_k^2)$, $D_\alpha = \text{diag}(\alpha_1, \dots, \alpha_k)$ and $(L, D_\kappa) \in \Theta_G^k$ and correspond to modified Cholesky decomposition of Q .

When skewness parameters are all zero, the density (4) reduces to the multivariate normal. Figure 2 shows contour plots of the SGD model ($k = 2$) for different skewness parameters. Although Definition 4 depends on the ordering of the vertices, this is not as restrictive as it first appears. The ordering is essentially another parameter to be specified and can be viewed as imposing extra information. In sequel, we show that the SGD model is supported by the property of conditional independence.

Theorem 1: Let \mathbf{X} be closed skew-normal distributed corresponding to Definition 4. Then for $i < j$, we have

$$X_i \perp X_j | \mathbf{x}_{-ij} \quad \Leftrightarrow \quad Q_{ij} = 0 \quad \Leftrightarrow \quad L_{ij} = 0 \quad \Leftrightarrow \quad (i, j) \notin E.$$

This is one of the main results. It simply says that similar to the Gaussian graphical models the nonzero pattern of Q determines G , so we can read off from Q

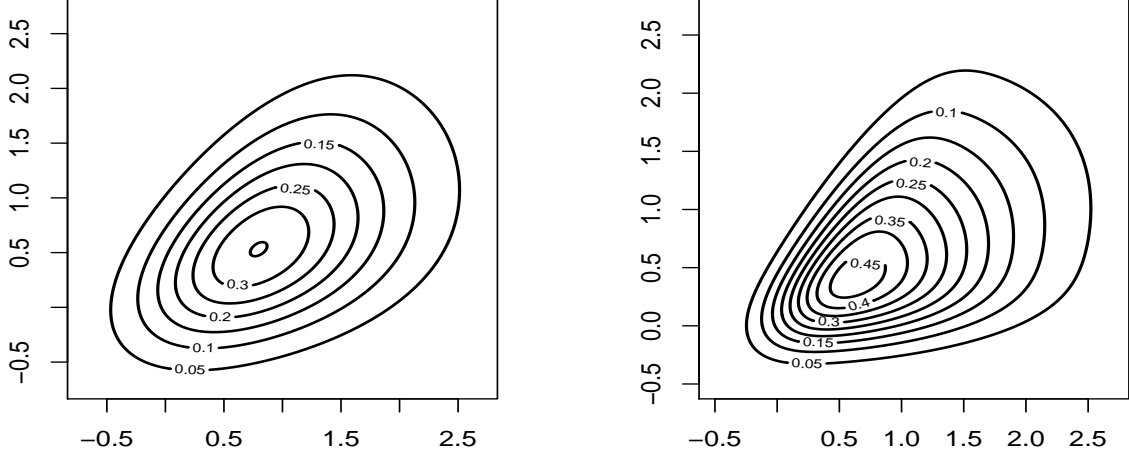


Figure 2: Left: contour plot of the SGD model with parameters $\boldsymbol{\mu} = (0, 0)'$, $D_\kappa = \text{diag}(1, 1)$, $L_{12} = -0.5$ and $\alpha_1 = \alpha_2 = 2$. Right: contour plot of the SGD model with parameters $\boldsymbol{\mu} = (0, 0)'$, $D_\kappa = \text{diag}(1, 1)$, $L_{12} = -0.5$ and $\alpha_1 = \alpha_2 = 4$.

whether X_i and X_j are conditionally independent, so the natural way to parametrize the SGD model is by its precision matrix Q . Although, the results in this article are based on decomposable graphical models, in sequel we have a theorem based on an arbitrary undirected graph.

Lemma 2: Consider an arbitrary undirected graph $G = (V, E)$ where $V = \{1, 2, \dots, k\}$. Let L be the modified cholesky decomposition of the precision matrix corresponding to graph G . Define the set $F(i, j) = \{i + 1, \dots, j - 1, j + 1, \dots, k\}$. Now, if $F(i, j)$ separates $i < j$ in G , then $L_{ij} = 0$ (see Rue and Held, 2005).

Theorem 2: Consider an arbitrary undirected graph $G = (V, E)$. Let L be the modified Cholesky triangle of precision matrix $Q \in P_G^k$. Suppose \mathbf{X} is closed skew-normal distributed (4), then for $i < j$, we have

$$\text{if } F(i, j) \text{ separates two vertices } i \text{ and } j \Rightarrow X_i \perp X_j | \mathbf{x}_{-ij}.$$

Alternatively, using the representation of Dominguez-Molina et al. (2003) for CSN families, a stochastic representation of the SGD model is

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + L^{-1} D_\kappa^{-\frac{1}{2}} D_\alpha (I + D_\alpha^2)^{-\frac{1}{2}} \mathbf{U} + L^{-1} D_\kappa^{-\frac{1}{2}} (I + D_\alpha^2)^{-\frac{1}{2}} \mathbf{V}, \quad (5)$$

where \mathbf{V} and \mathbf{U} are independent multivariate normal distribution $N_k(\mathbf{0}, I_k)$ and a multivariate half-normal distribution $HN_k(\mathbf{0}, I_k)$, respectively. From a computational point of view, this representation is useful because it implies that a random vector distributed according to (4) can be generated using two independent normal

random vectors. The mean vector and covariance matrix of \mathbf{X} is

$$\begin{aligned} E(\mathbf{X}) &= \boldsymbol{\mu} + L^{-1} D_{\kappa}^{-\frac{1}{2}} D, \\ Cov(\mathbf{X}) &= (L' D_{\kappa}^{\frac{1}{2}} (I_k - D^2) D_{\kappa}^{\frac{1}{2}} L)^{-1} \end{aligned}$$

where diagonal matrix D is $D = \sqrt{\frac{2}{\pi}} D_{\alpha} (I_k + D_{\alpha}^2)^{-\frac{1}{2}}$. Hence, the zero elements of inverse of covariance matrix \mathbf{X} are the same of those of the precision matrix Q .

4 Bayesian analysis

In this section, we will discuss Bayesian inference of the SGD model (4). Assume n independent observations $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ from this model. We use the representation (5) and the following hierarchical representation

$$\begin{aligned} \mathbf{X}_j | \mathbf{U}_j &\sim N_k(\boldsymbol{\mu} + L^{-1} D_{\delta} \mathbf{U}_j, (L' D_{\omega} L)^{-1}), \quad \boldsymbol{\delta}, \boldsymbol{\mu} \in \mathbb{R}^k, \quad (L, D_{\omega}) \in \Theta_G \\ \mathbf{U}_j &\sim HN_k(\mathbf{0}, I_k), \quad j = 1, \dots, n, \end{aligned} \quad (6)$$

where $D_{\omega} = diag(\omega_1^2, \dots, \omega_k^2)$ and $D_{\delta} = diag(\delta_1, \dots, \delta_k)$ with $\delta_i = \frac{\alpha_i}{\kappa_i \sqrt{1 + \alpha_i^2}}$ and $\omega_i = \kappa_i \sqrt{1 + \alpha_i^2}$. The later reparametrization has been imposed to ease the computations. We will now discuss the prior distribution for the unknown parameters $\boldsymbol{\eta} = (\boldsymbol{\mu}, \boldsymbol{\delta}, L, D_{\omega})$. With regard to relation between δ_i and ω_i in the mentioned reparametrization (i.e. $\delta_i = \frac{\alpha_i}{\omega_i}$), we will use a correlated prior for $\boldsymbol{\delta} = (\delta_1, \dots, \delta_k)$ and $\boldsymbol{\omega} = (\omega_1, \dots, \omega_k)$;

$$\pi(\boldsymbol{\delta} | \boldsymbol{\omega}) = N_k(\mathbf{0}, b_1 D_{\omega}^{-1}),$$

where $b_1 \in \mathbb{R}_+$ is fixed. For the priors on $\boldsymbol{\mu} \in \mathbb{R}^k$ and $(L, D_{\omega}) \in \Theta_G$. We will discuss two separate prior structures corresponding to independent proper and noninformative priors for these parameters.

Independent proper priors on $\boldsymbol{\mu}$, $\boldsymbol{\omega}^2$ and L : We take normal priors for mean vector and lower triangular matrix and gamma prior for precision parameters as follows

$$\begin{aligned} \pi(\boldsymbol{\mu}) &= N_k(\boldsymbol{\mu}_0, b_2 I_k), \\ \pi(\boldsymbol{\omega}^2) &= \prod_{i=1}^k G(b_3, b_4), \\ \pi(L_{1.}^{\neq 0}, \dots, L_{k-1.}^{\neq 0}) &= \prod_{i=1}^{k-1} N_{\|N^{\prec(i)}\|}(\mathbf{0}, b_5 I_{\|N^{\prec(i)}\|}), \end{aligned} \quad (7)$$

where $G(a, b)$ denotes the gamma distribution, $L_i^{\neq 0} = (L_{ij})_{\{i < j, (i, j) \in E\}}$ denotes the nonzero off-diagonal elements of i 'th row of L and $N^{\prec(i)} := \{j : (i, j) \in E, i < j\}$. Also, $\boldsymbol{\mu}_0 \in \mathbb{R}^k$ and $b_i \in \mathbb{R}_+, i = 2, 3, 4, 5$, are known hyperparameters.

Non-informative priors on $\boldsymbol{\mu}$, $\boldsymbol{\omega}^2$ and L : Assign a common noninformative prior distributions for the mean vector of the form $\pi(\boldsymbol{\mu}) \propto 1$. Khare and Rajaratnam

(2011) formed a rich and flexible class of Wishart priors for decomposable covariance graphs in Gaussian covariance graph models in which marginal independence among the components of a multivariate random vector is encoded by means of a graph G . They also considered the case where G is decomposable. Although covariance graph models are distinctly different from the traditional concentration graph models, after some modifications, their prior is extendable in our problem case. In this context, the class of measures on $(L, D_\omega) \in \Theta_G$ is provided with density

$$\pi(L \in dL, D_\omega \in dD_\omega) \propto \prod_{i=1}^k (\omega_i^2)^{\psi_i/2-1} e^{-\frac{1}{2}\text{tr}((L'D_\omega L)\Psi)} \prod_{i < j, (i,j) \in E} dL_{ij} \prod_{i=1}^k d\omega_i^2, \quad (8)$$

where positive definite matrix Ψ and $\psi_i \in \mathbb{R}_+$, $i = 1, \dots, k$ are known hyperparameters. Theorem 3 provides a sufficient condition for the existence of a normalizing constant for $\pi(L, D_\omega)$.

Theorem 3: Let $dL := \prod_{i < j, (i,j) \in E} dL_{ij}$ and $dD_\omega := \prod_{i=1}^k d\omega_i^2$. if $\psi_i > ||N^\prec(i)||$, then

$$\int_{\Theta_G} \prod_{i=1}^k (\omega_i^2)^{\psi_i/2-1} \exp\{-\frac{1}{2}\text{tr}((L'D_\omega L)\Psi)\} dD_\omega dL < \infty. \quad (9)$$

A standard noninformative prior for lower triangular matrix L and precision parameters ω^2 can be chosen by respecting the zeros for Ψ and ψ_i , $i = 1, \dots, k$. Hence, the resulting non-informative prior on μ , L and ω^2 is of the form

$$\pi(\mu, L, \omega^2) \propto \prod_{i=1}^k \frac{1}{\phi 2_i}. \quad (10)$$

Theorem 4 shows that the posterior distribution is proper under prior distribution (10). To show this, we use a result by Mouchart (1976) and Florens *et al.* (1990) which implies that the posterior distribution exists as a proper only when

$$p(\mathbf{x}) = \int p(\mathbf{x}|\mu, \omega^2, \delta, L) \pi(\mu, \omega^2, \delta, L) d\mu d\omega^2 d\delta dL < \infty.$$

Theorem 4: Under the standard noninformative prior in (10) and with n independent replication from the hierarchical model (6), the posterior distribution of parameters exists if $n \geq \max\{||N^\prec(i)||\} + 2$.

We will now discuss how to generate samples from the posterior distribution. To facilitate the sampling, we introduce the latent variables $\mathbf{U} = (\mathbf{U}_1, \dots, \mathbf{U}_n)$, and then use Gibbs sampling to generate samples. The block full conditionals are as

Table 1: The value of hyperparameters under different prior distributions.

Hyperparameter	Prior (7)	Prior (8)	Prior (10)
v_μ	$1/b_2$	\dots	0
s_{ω_i}	b_3	$\psi_i/2$	0
r_{ω_i}	b_4	$\frac{1}{2}L_i.\Psi L'_i.$	0
V_{L_i}	$1/b_5 I_k$	$\omega_i^2 \Psi$	0

follows:

$$\begin{aligned}
\pi(\mathbf{u}|\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\delta}, \omega^2, L) &= \prod_{i=1}^n \prod_{j=1}^k HN\left(\frac{\omega_j^2 \delta_j}{1 + \omega_j^2 \delta_j^2} L_{j.}(\mathbf{x}_i - \boldsymbol{\mu}), \frac{1}{1 + \omega_j^2 \delta_j^2}\right), \\
\pi(\boldsymbol{\delta}|\mathbf{x}, \mathbf{u}, \boldsymbol{\mu}, \omega^2, L) &= N_k(\Sigma_\delta^{-1}[\Sigma_{i=1}^n D_{u_i} L(\mathbf{x}_i - \boldsymbol{\mu})], \Sigma_\delta^{-1} D_\omega^{-1}), \\
\pi(\boldsymbol{\mu}|\mathbf{x}, \mathbf{u}, \boldsymbol{\delta}, \omega^2, L) &= N_k(\Sigma_\mu^{-1}[(L' D_\omega L) \Sigma_{i=1}^n (\mathbf{x}_i - L^{-1} D_\delta \mathbf{u}_i) + v_\mu \boldsymbol{\mu}_0], \Sigma_\mu^{-1}), \\
\pi(\omega^2|\mathbf{x}, \mathbf{u}, \boldsymbol{\mu}, \boldsymbol{\delta}, L) &= \prod_{i=1}^k G(s_{\omega_i} + \frac{n+1}{2}, r_{\omega_i} + \frac{1}{2} L_i. S_u L'_i. + \frac{1}{2b_1} \delta_i^2),
\end{aligned}$$

where $D_{u_i} = \text{diag}(\mathbf{u}_i)$, $i = 1, \dots, n$, and

$$\begin{aligned}
\Sigma_\delta &= \Sigma_{i=1}^n D_{u_i}^2 + 1/b_1 I_k, \quad \Sigma_\mu = n(L' D_\omega L) + v_\mu I_k, \\
S_u &= \Sigma_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu} - L^{-1} D_\delta \mathbf{u}_i)(\mathbf{x}_i - \boldsymbol{\mu} - L^{-1} D_\delta \mathbf{u}_i)'.
\end{aligned}$$

The full conditional for $L_i^{\neq 0}$, $i = 1, \dots, k-1$, is

$$\pi(L_i^{\neq 0} | L \setminus L_i^{\neq 0}, \mathbf{x}, \mathbf{u}, \boldsymbol{\mu}, \boldsymbol{\delta}, \omega^2) = N_{||N \prec (i)||}(\Sigma_{(ii)}^{-1}[\omega_i^2 \delta_i M_i. - \boldsymbol{\zeta}_{(i)}], \Sigma_{(ii)}^{-1}),$$

where $M_i. = (M_{ij})_{\{i < j, (i,j) \in E\}}$ with $M = \sum_{i=1}^n \mathbf{u}_i \mathbf{x}_i'$. Additionally, $\Sigma_{(ii)}$ and $\boldsymbol{\zeta}_{(i)}$ are submatrices of $\Sigma^{(i)} = \omega_i^2 S + V_{L_i}$ with $S = \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})'$ such that $\boldsymbol{\zeta}_{(i)} = (\Sigma_{ij}^{(i)})_{\{i < j, (i,j) \in E\}}$ and $\Sigma_{(ii)} = (\Sigma_{kl}^{(i)})_{\{(k,l) | (k,l) \in E, k,l > i\}}$. Note that the resulting posterior distributions under different values for hyperparameters can be determined based on Table 1.

5 Simulation study

The SGD model introduces the extra parameter $\boldsymbol{\delta}$ beyond the parametrization of the usual Gaussian graphical model, so now we want to examine to what extent information on this parameter can be recovered from data. We assign a diffuse prior on $\boldsymbol{\delta}$ using $b_1 = 100$, and standard noninformative priors (10) for the other parameters. We generate data from the SGD model (6) with $k = 3$ variables under the neighborhood graph (3). We use a sample size of $n = 200$ with $\boldsymbol{\mu} = 5\mathbf{1}_3$ and $D_\omega = I_3$. We focus on the skewness parameters δ_1 , δ_2 and δ_3 and two nonzero elements L_{12} and L_{23} of lower triangular matrix L as inference is most challenging for these parameters. The data sets based on different values for these parameters has been determined as follows:

Case A: Four data sets generated by $L_{12} = L_{23} = -0.5$ and with four different values of skewness parameters given by $\delta_1 = \delta_2 = \delta_3 = -1, 1, 2, 3$

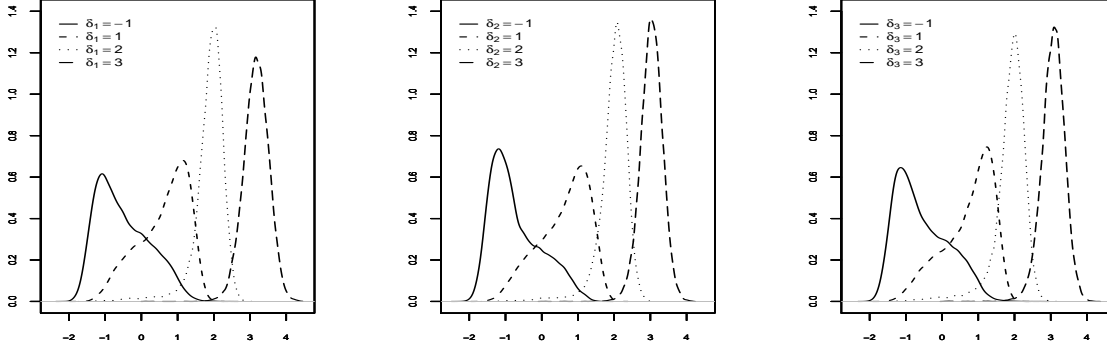


Figure 3: Posterior distributions for δ_1 , δ_2 and δ_3 under four simulated data sets in Case A. The legends indicate the values of δ_i 's used to generate data sets.

Case B: Four data sets generated by $\delta_1 = \delta_2 = \delta_3 = 2$ and $L_{12} = L_{12} = -1, -0.5, 0.5, 1$

Case C: One data set generated with $\delta_1 = 3$, $\delta_2 = -2$, $\delta_3 = -4$, $L_{12} = -0.5$ and $L_{23} = 0.5$.

The motivation for choosing the parameters in case C is that for a given observation vector $\mathbf{X} = (X_1, X_2, X_3)$, we have

$$\begin{aligned} X_1 &\stackrel{d}{=} 3|U_1| - |U_2| + |U_3| + V_1 + 0.5V_2 - 0.25V_3, \\ X_2 &\stackrel{d}{=} -2|U_2| + 2|U_3| + V_2 - 0.5V_3, \\ X_3 &\stackrel{d}{=} -4|U_3| + V_3. \end{aligned}$$

where $\mathbf{U} = (U_1, U_2, U_3)'$, $\mathbf{V} = (V_1, V_2, V_3)' \stackrel{iid}{\sim} N_3(\mathbf{0}, I_3)$. Hence, we can easily see that $(X_1, X_2, X_3)' \stackrel{d}{=} (X_1, -X_2, X_3)'$, so the marginal density for X_2 is symmetric.

Our results are based on a MCMC chain of length 500000 with a burn-in of 100000, using the block MCMC algorithm in Section 5. Figure 3 displays the posterior distributions for δ_1 , δ_2 and δ_3 under four data sets introduced in Case A. These figures clearly indicate that the data allow for meaningful inference on $\boldsymbol{\delta}$ since the posterior distributions assign a large mass to neighborhoods of the values used to generate the data, specially for larger values of the skewness. Figure 4 displays the posterior inference for L_{12} and L_{23} under four data sets in Case B. Figure 5 displays the posterior distributions for (left figure) δ_1 , δ_2 , δ_3 and (right figure) L_{12} and L_{23} under a data set introduced in Case C which indicates the data clearly allows for good inference.

6 Case studies

In this section, we apply our approach to two data sets: student's mathematics marks from Mardia et al. (1979) and the carcass data from Busk et al. (1999). We compare the results with those obtained from the Gaussian model. Our results are based on a MCMC chain of length 700000 with a burn-in 200000. All results in this section were computed under prior (7) with $b_1 = 100$, $\boldsymbol{\mu}_0 = \mathbf{0}$, $b_2 = 10^4$, $b_3 = b_4 = 10^{-6}$ and $b_5 = 100$ which gives us a diffuse prior.

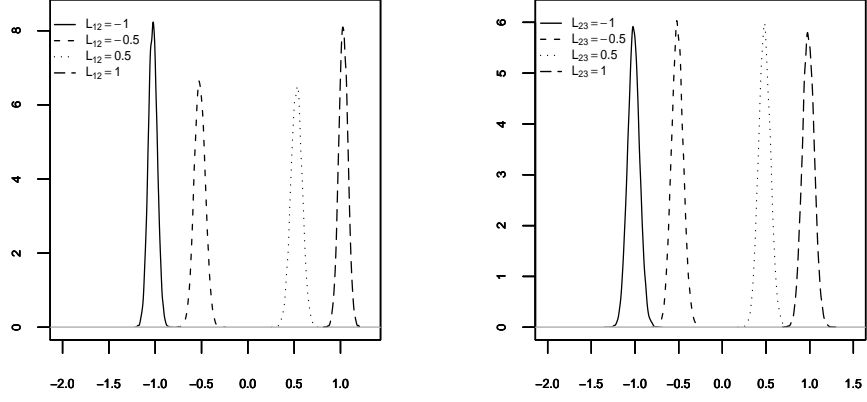


Figure 4: Posterior distributions for L_{12} and L_{23} under four simulated data sets in Case B. The legends indicate the values of L_{12} and L_{23} used to generate the data sets.

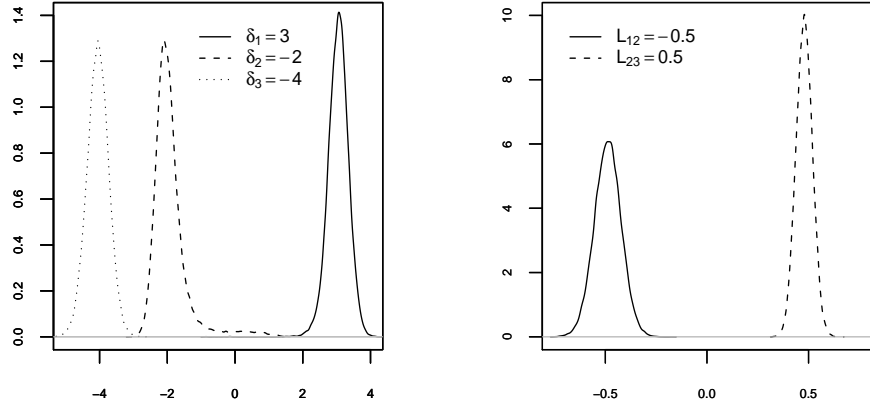


Figure 5: Posterior distributions for (left figure) δ_1 , δ_2 , δ_3 and (right figure) L_{12} and L_{23} under simulated data in Case C. The legends indicate the values of these parameters used to generate the data.

Table 2: The sample partial correlation matrix of the mathematics marks.

	Mechanics	Vectors	Algebra	Analysis	Statistics
Mechanics	1	0.33	0.23	0.00	0.02
Vectors		1	0.28	0.08	0.02
Algebra			1	0.43	0.36
Analysis				1	0.25
Statistics					1

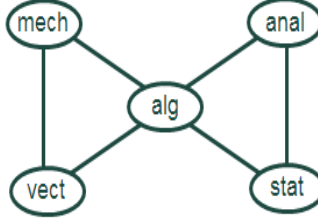


Figure 6: Neighborhood graph of mathematics marks.

6.1 Mathematics Marks

This data set come from Mardia et al. (1979), and consists of examination marks of 88 students in the five subjects mechanics, vectors, algebra, analysis and statistics. Mechanics and vectors were closed book examinations and the reminder were open book. All variables are measured on the same scale (0-100). Table 2 displays sample partial correlation matrix between these variables. The element in the upper righthand block are all near zero which it means that we can consider mechanics and analysis conditionally independent on the other remaining variables, as are mechanics and statistics, vectors and statistics and finally vectors and analysis. These assumptions have been described in neighborhood graph in Figure 6 as suggested by Whittaker (1990). For exploratory purpose, the histograms of variables are plotted in Figure 7. These histograms suggest that analysis and statistics marks have skewed distributions in left and right, respectively. We used the following ordering for our analysis: {Mechanics, Vectors, Algebra, Analysis, Statistics}. The posterior mean (standard deviation) estimates of nonzero elements of upper triangular matrix L under GG and SGDG model are as follows

$$\hat{L}_{GG} = \begin{pmatrix} 1 & -0.46_{(0.15)} & -0.55_{(0.18)} & 0.00 & 0.00 \\ & 1 & -0.75_{(0.11)} & 0.00 & 0.00 \\ & & 1 & -0.35_{(0.06)} & -0.23_{(0.05)} \\ & & & 1 & -0.52_{(0.07)} \\ & & & & 1 \end{pmatrix},$$

$$\hat{L}_{SGDG} = \begin{pmatrix} 1 & -0.46_{(0.14)} & -0.55_{(0.17)} & 0.00 & 0.00 \\ & 1 & -0.76_{(0.09)} & 0.00 & 0.00 \\ & & 1 & -0.32_{(0.06)} & -0.24_{(0.05)} \\ & & & 1 & -0.44_{(0.05)} \\ & & & & 1 \end{pmatrix}.$$

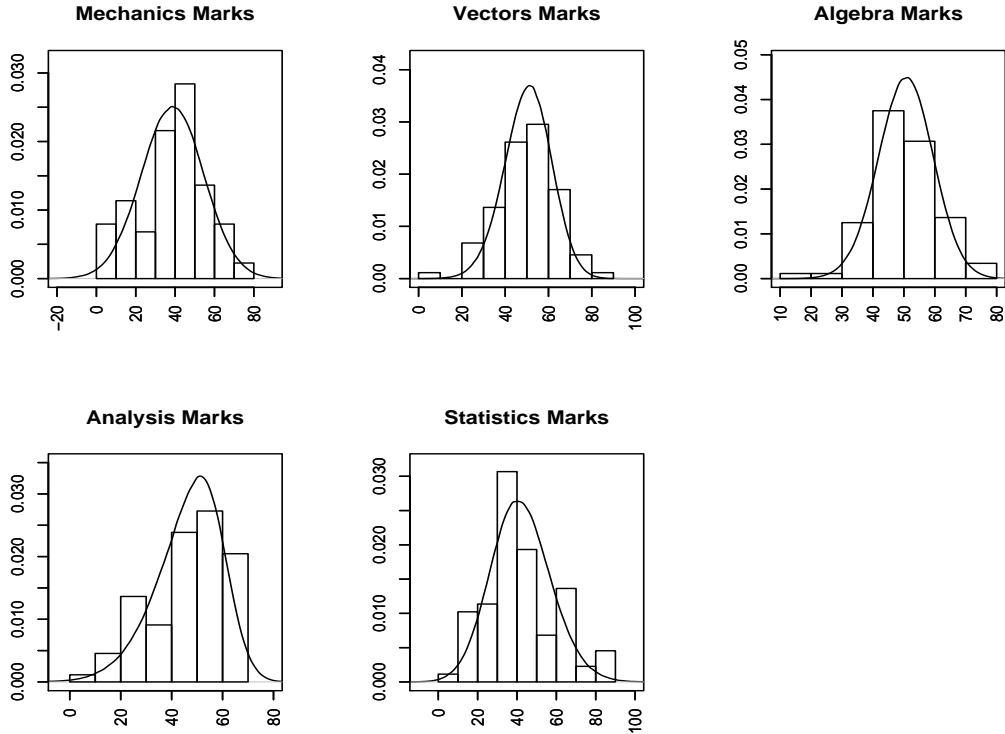


Figure 7: Histograms of mathematics marks and fitted models to those.

The posterior mean estimates of the other parameters is shown in Table 3. Recall that δ_i is not skewness parameter of i th variable only, but the δ_i 's in Table 3 can be approximately interpreted as skewness parameters of the rows of $\hat{L}_{SGDG}X$ where $X=(\text{Mechanics, Vectors, Algebra, Analysis, Statistics})$. Figure 7 shows the histograms of variables with the fitted SGD model. The fits seems adequate.

To compare the GG and SGD model, we computed the Bayes factor using the modified harmonic mean estimator \hat{p}_4 of Newton and Raftery (1994). The Bayes factor in favor of the SGD model is 3×10^{37} , which indicates overwhelming support for the SGD model. We also tried other vertex-orderings corresponding to a perfect vertex elimination scheme, but the chosen ordering in the first of this section (i.e. $\{\text{Mechanics, Vectors, Algebra, Analysis, Statistics}\}$) was supported by the Bayes factor. Additionally, note that although the SGD model depends on the ordering of the vertices, it give a better fit in compare with the corresponding GG model over all vertex-orderings corresponding to a perfect vertex elimination scheme.

6.2 Carcass Data

The carcass data from gRbase package of R contains measurements of the thickness of meat and fat layers together with the lean meat percentage of 344 slaughter carcasses at three Danish slaughter houses. Seven variables has been defined in this data set as follows:

- F11, F12, F13: Thickness of fat layer at 3 different locations on the back of the carcass.

Table 3: Mathematics marks: Posterior means (standard deviation) for parameters.

i	SGDG			GG	
	β	ω^2	δ	β	ω^2
1	45.83 _(5.75)	0.0057 _(0.0015)	0.38 _(6.15)	39.16 _(1.83)	0.0051 _(0.0008)
2	61.21 _(4.83)	0.0235 _(0.0174)	-9.17 _(5.63)	50.68 _(1.44)	0.0091 _(0.0014)
3	55.18 _(3.27)	0.0469 _(0.0279)	-5.51 _(4.15)	50.64 _(1.12)	0.0211 _(0.0032)
4	56.17 _(2.30)	0.2131 _(0.3696)	-18.29 _(1.88)	46.77 _(1.56)	0.0071 _(0.0011)
5	30.53 _(3.68)	0.0065 _(0.0025)	14.54 _(3.89)	42.25 _(1.88)	0.0034 _(0.0005)

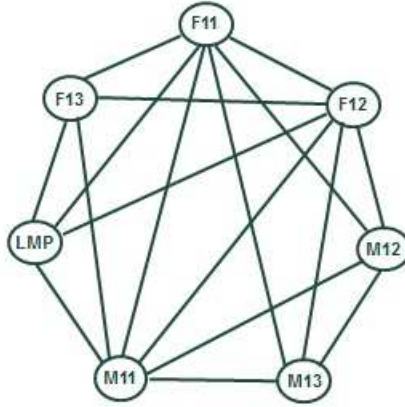


Figure 8: Neighborhood graph of carcass data.

- M11, M12, M13: Thickness of meat layer at 3 different locations on the back of the carcass.
- LMP: Lean meat percentage determined by dissection

This data set has been used for estimating the parameters in a prediction formula for prediction of lean meat percentage on the basis of the thickness measurements on the carcass. Data are described in detail in Busk et al. (1999). Hojsgaard et al. (2012) provided some neighborhood graphs for these variables in term of different model selection methods. Based on the BIC criterion, they proposed the neighborhood graph in Figure 8.

We used first the ordering: $\{M12, M13, M11, LMP, F13, F11, F12\}$. A summary of the posterior inference on the parameters in the models is provided in Table 4. We also present two quantities that are directly comparable between the SGD and GG model. In Figure 9, we have displayed the fitted SGD model to this data set. The SGD model seems to fit well and includes skewness for LMP, F13, F11 and F12. The Bayes factor in favor of the SGD versus the GG model was estimated to 1.7×10^8 . We also tried an alternative orderings for our analysis, but they gave all a worse fit.

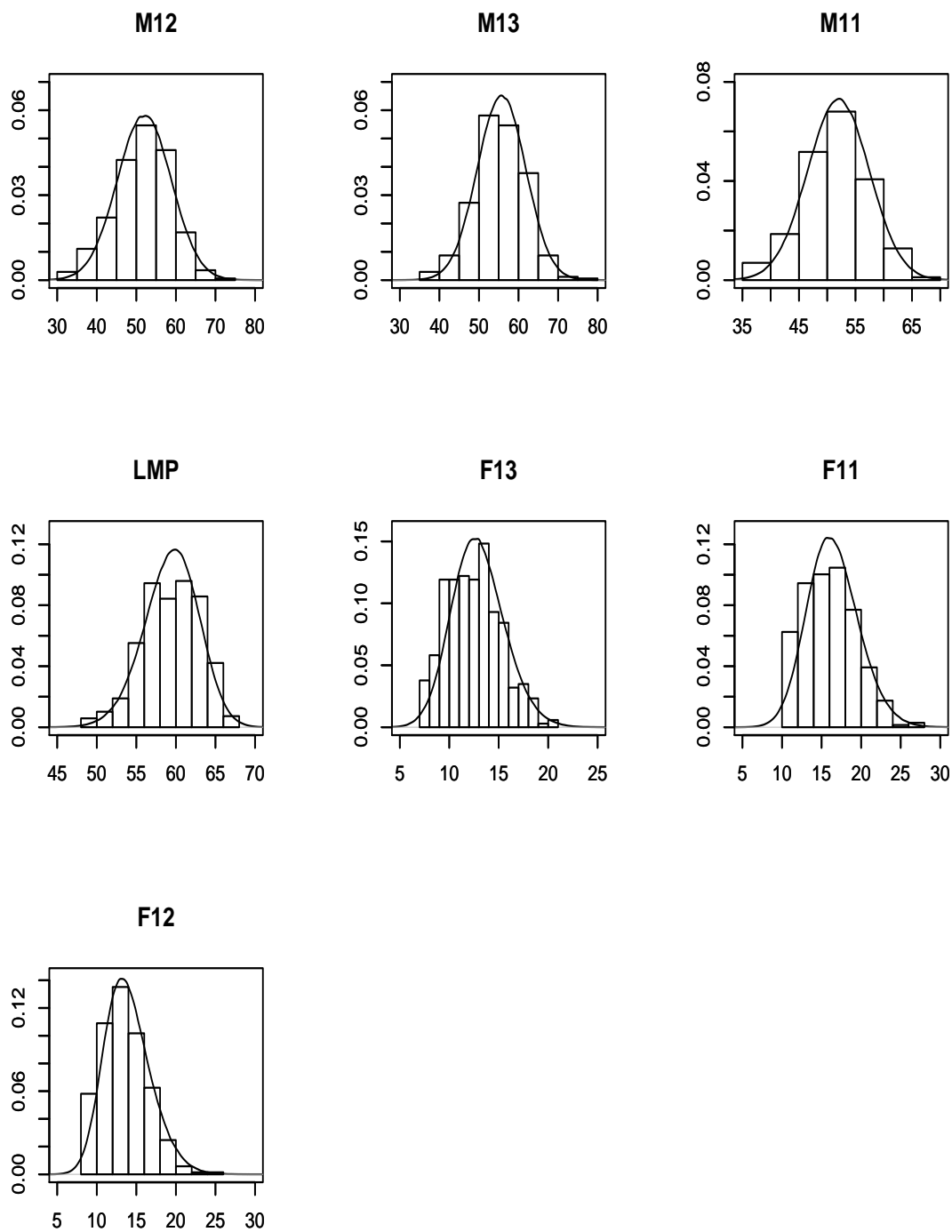


Figure 9: Histograms of carcass data and fitted models to those.

Table 4: Carcass Data: Posterior means (standard deviation) for parameters.

i	SGDG				GG	
	β	ω^2	δ		β	ω^2
1	51.67 _(2.62)	0.152 _(1.41)	0.54 _(0.025)		52.01 _(0.41)	0.133 _(0.010)
2	55.89 _(2.38)	0.112 _(1.70)	0.87 _(0.021)		55.70 _(0.35)	0.096 _(0.007)
3	52.44 _(2.15)	0.047 _(2.46)	1.32 _(0.008)		51.98 _(0.31)	0.041 _(0.003)
4	63.81 _(0.96)	0.326 _(0.91)	-2.06 _(0.085)		59.38 _(0.19)	0.196 _(0.015)
5	10.24 _(0.83)	0.615 _(0.79)	0.43 _(0.131)		12.95 _(0.15)	0.510 _(0.038)
6	13.46 _(0.98)	0.342 _(1.06)	0.25 _(0.062)		16.49 _(0.18)	0.293 _(0.023)
7	10.96 _(0.54)	0.286 _(0.66)	3.75 _(0.090)		13.96 _(0.16)	0.112 _(0.009)

7 Conclusions

In this paper, we have proposed a novel decomposable graphical model to accommodate skew-Gaussian decomposable graphical model (SGDG). The SGDGD model reflect conditional independencies among the components of the multivariate closed skew normal random vector with respect to a decomposable graph, and includes the Gaussian graphical models as a particular case.

We develop a family of conjugate prior distributions for precision matrix, and derive a condition which ensures propriety of the posterior distribution corresponding to the standard noninformative priors for mean vector, precision parameters and lower triangular matrix as well as proper prior for skewness parameter. We successfully apply our new model to two data sets, showing great improvement on the Gaussian graphical model.

Since the SGDGD model satisfy the conditional independence property only under a decomposable graph, so an extension of the SGDGD model for other graphs requires further research.

APPENDIX: PROOFS

To prove Theorem 1, we use the following factorization criterion.

Lemma 4: Two random variables X and Y are called conditionally independent given Z iff there exist some functions f and g such that $\pi(x, y, z) = f(x, z)g(y, z)$ (see Rue and Held (2005)).

Proof of Theorem 1: We partition \mathbf{X} as $(X_i, X_j, \mathbf{X}_{-ij})$ and then use the multivariate version of the factorization criterion on $\pi(x_i, x_j, \mathbf{x}_{-ij})$. Fix $i \neq j$ and assume $\boldsymbol{\mu} = \mathbf{0}$ without loss of generality. From (4) we get

$$\pi(x_i, x_j, \mathbf{x}_{-ij}) \propto \exp\left(-\frac{1}{2} \sum_{m,l} x_m Q_{ml} x_l\right) \prod_{r=1}^k \Phi_1(\alpha_r \kappa_r L'_r \mathbf{x}; 0, 1),$$

where κ_r is root square of the r 'th diagonal element of D_κ and L_r is r 'th row of L . Let $i < j$ without loss of generality. Now, we define set

$$I = \{r : L_{ri} \neq 0 \text{ and } L_{rj} = 0 \text{ for } r \leq i\}$$

Then, due to the ordering of the vertices corresponds to a perfect vertex elimination scheme and with respect to Definition 2, we know if $v < i < j$ then the neighbors of vertex v can not be vertex i and j together if and only if vertices i and j are not neighbors. Hence, we have

$$\begin{aligned} \pi(x_i, x_j, \mathbf{x}_{-ij}) &\propto \exp\left(-\frac{1}{2} x_i x_j (Q_{ij} + Q_{ji}) - \frac{1}{2} \sum_{\{m,l\} \neq \{i,j\}} x_m Q_{ml} x_l\right) \\ &\times \prod_{r \in I} \Phi_1(\alpha_r \kappa_r L'_r \mathbf{x}; 0, 1) \prod_{r \notin I} \Phi_1(\alpha_r \kappa_r L'_r \mathbf{x}; 0, 1). \end{aligned}$$

Second term does not involve $x_i x_j$ while first term involves $x_i x_j$ iff $Q_{ij} \neq 0$. Thus, one example for functions f and g in lemma can be defined as

$$\begin{aligned} f(x_i, \mathbf{x}_{-ij}) &= \exp\left(-\frac{1}{2} \sum_{\{m,l\} \neq j} x_m Q_{ml} x_l\right) \prod_{r \in I} \Phi_1(\alpha_r \kappa_r L'_r \mathbf{x}; 0, 1) \\ g(x_j, \mathbf{x}_{-ij}) &= \exp\left(-\frac{1}{2} \sum_{\{m,l\} \neq i} x_m Q_{ml} x_l\right) \prod_{r \notin I} \Phi_1(\alpha_r \kappa_r L'_r \mathbf{x}; 0, 1). \end{aligned}$$

iff $Q_{ij} \neq 0$. The claim then follows. ■.

Proof of Theorem 2: At first, we recall if G be a nondecomposable undirected graph, then it is possible that the zero elements of precision matrix Q are not reflected in its modified cholesky decomposition L . To prove this theorem, we proceed similar to Theorem 1. Fix $i < j$ and assume $\boldsymbol{\mu} = \mathbf{0}$ without loss of generality. Here, we know when $F(i, j)$ separates vertices i and j , then $L_{ij} = 0$ and it is not possible to have a pass from i to j with passing from some vertex that the number of all

of them are smaller than i . Thus, if $v < i < j$, then v can not be the neighbors of vertex i and j together. Hence, we can write $\pi(x_i, x_j, \mathbf{x}_{-ij})$ similar to (11) with same definition for set I . The rest of the proof is similar to that of Theorem 1. ■

Proof of Theorem 3: At first, we denote $L_{i\cdot}$ as i 'th row of L . Now, we have

$$\begin{aligned} & \int \prod_{i=1}^k (\omega_i^2)^{\psi_i/2-1} \exp\{-\frac{1}{2} \text{tr}((L' D_\omega L) \Psi)\} dD dL \\ &= \int \prod_{i=1}^k (\omega_i^2)^{\psi_i/2-1} \exp\{-\frac{1}{2} (L \Psi L')_{ii} \omega_i^2\} dD dL \propto \int \prod_{i=1}^k \frac{1}{(L \Psi L')_{ii}^{\psi_i/2}} dL \\ &\propto \prod_{i=1}^{k-1} \int \frac{1}{(L_{i\cdot} \Psi L_{i\cdot}')^{\psi_i/2}} dL_{i\cdot} = \prod_{i=1}^{k-1} \int_{\mathbb{R}^{|N^{\prec(i)}|}} \frac{1}{(L_{i\cdot}^* \Psi^{(i)} L_{i\cdot}^{*'})^{\psi_i/2}} dL_{i\cdot}^{\neq 0} \end{aligned}$$

where $L_{i\cdot}^* = (1 \ L_{i\cdot}^{\neq 0})'$. Also, $\Psi^{(i)}$ is a submatrix of Ψ corresponding to the elements of $L_{i\cdot}^*$. We are now in the same line with Khare and Rajaratnam (2011) and similarly, we can show that the integral is finite if $\psi_i > ||N^{\prec(i)}||$. ■

Proof of Theorem 4: Under the introduced reparemeterizations in hierarchical model (6), the conditional density of \mathbf{X} becomes

$$p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\omega}^2, \boldsymbol{\delta}, L) = \left(\frac{2}{\pi}\right)^{k/2} |D_g|^{\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})' Q_g (\mathbf{x}-\boldsymbol{\mu})} \prod_{i=1}^k \Phi(\delta_i \omega_i \sqrt{g(\omega_i, \delta_i)} L(\mathbf{x} - \boldsymbol{\mu}); 0, 1),$$

where $g(\omega_i, \delta_i) = \frac{\sigma_{2i}^2}{1+\delta_i^2 \sigma_{2i}^2}$, $D_g = \text{diag}(g(\omega_1, \delta_1), \dots, g(\omega_k, \delta_k))$ and $Q_g = L' D_g L$. Now, since $0 \leq \Phi(\cdot) \leq 1$, we have

$$\begin{aligned} p(\mathbf{x}) &= \int_{\mathbb{R}^N} \int_{\mathbb{R}^k} \int_{\mathbb{R}_+^k} \int_{\mathbb{R}^k} p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\omega}^2, \boldsymbol{\delta}, L) \pi(\boldsymbol{\mu}) \pi(\boldsymbol{\delta}|\boldsymbol{\omega}^2) \pi(L, D_\omega) d\boldsymbol{\mu} d\boldsymbol{\omega}^2 d\boldsymbol{\delta} dL \\ &< \int_{\mathbb{R}^N} \int_{\mathbb{R}^k} \int_{\mathbb{R}_+^k} \int_{\mathbb{R}^k} |D_g|^{\frac{n}{2}} e^{-\frac{1}{2} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})' Q_g (\mathbf{x}_i - \boldsymbol{\mu})} \pi(\boldsymbol{\delta}|\boldsymbol{\omega}^2) \pi(L, D_\omega) d\boldsymbol{\mu} d\boldsymbol{\omega}^2 d\boldsymbol{\delta} dL \\ &= K_1 \int_{\mathbb{R}^k} f_N^k(\boldsymbol{\mu}|\bar{\mathbf{x}}, (nQ_g)^{-1}) d\boldsymbol{\mu} \\ &\times \int_{\mathbb{R}^N} \int_{\mathbb{R}^k} \int_{\mathbb{R}_+^k} |D_g|^{\frac{n}{2}} |Q_g|^{-\frac{1}{2}} e^{-\frac{1}{2} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})' Q_g (\mathbf{x}_i - \bar{\mathbf{x}})} \pi(\boldsymbol{\delta}|\boldsymbol{\omega}^2) \pi(L, D_\omega) d\boldsymbol{\omega}^2 d\boldsymbol{\delta} dL \end{aligned}$$

where $N = \sum_{i=1}^{k-1} ||N^{\prec(i)}||$, $\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$ and K_1 is a constant. If we define $S = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$ and using variable transformation $\lambda_i = \delta_i \omega_i$, $i = 1, \dots, k$,

under $n \geq 2$, we get

$$\begin{aligned}
p(\mathbf{x}) &< K_2 \int_{\mathbb{R}^N} \int_{\mathbb{R}^k} \int_{\mathbb{R}_+^k} \prod_{i=1}^k \left(\frac{1}{1 + \delta_i^2 \phi 2_i} \right)^{\frac{n-1}{2}} \prod_{i=1}^k (\phi 2_i)^{\frac{n}{2}-1} e^{-\frac{1}{2} \text{tr}(Q_g S)} e^{-\frac{1}{2b_2} \sum_{i=1}^k \phi 2_i \delta_i^2} d\omega^2 d\delta dL \\
&= K_2 \int_{\mathbb{R}^N} \int_{\mathbb{R}^k} \int_{\mathbb{R}_+^k} \prod_{i=1}^k \left\{ \left(\frac{1}{1 + \lambda_i^2} \right)^{\frac{n-1}{2}} e^{-\frac{1}{2b_2} \lambda_i^2} (\phi 2_i)^{\frac{n-1}{2}-1} e^{-\frac{1}{2} \frac{(LSL')_{ii} \omega_i^2}{1 + \lambda_i^2}} \right\} d\omega^2 d\lambda dL \\
&\propto K_2 \int_{\mathbb{R}^N} \prod_{i=1}^k \frac{1}{(LSL')_{ii}^{\frac{n-1}{2}}} dL = \prod_{i=1}^{k-1} \int_{\mathbb{R}^{|N^{\prec(i)}|}} \frac{1}{(L_{i\cdot}^* \Psi^{(i)} L_{i\cdot}^{*'})^{\frac{n-1}{2}}} dL_{i\cdot}^{\neq 0}
\end{aligned}$$

where K_2 is another constant. Similarly to the proof of Theorem 3, we can show that the integral is finite if $n \geq \max\{||N^{\prec(i)}||\} + 2$. ■

References

- [1] Allard, D., Naveau, P. 2007. A new spatial skew-normal random field model. *Communications in Statistics* 36, 1821-1834.
- [2] Azzalini, A. 2005. The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics* 32, 159-188.
- [3] Azzalini, A., Capitanio, A. 1999. Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Association* 61, 579-602.
- [4] Busk, H., Olsen, E.V., Brondum, J. 1999. Determination of lean meat in pig carcasses with the Autofom classification system. *Meat Science* 52, 307-314.
- [5] Capitanio, A., Azzalini, A., Stanghellini, E. 2003. Graphical models for skew-normal variates. *Scandinavian Journal of Statistics* 30, 129-144.
- [6] Dominguez-Molina, J., Gonzalez-Farias, G., Gupta, A.K. 2003. The multivariate closed skew-normal distribution. Technical Report 03-12. Department of Mathematics and Statistics, Bowling Green State University.
- [7] Florens, J.P., Mouchart, M., Rolin, J.M. 1990. Invariance arguments in Bayesian statistics. In *Economic Decision Making: Games, Econometrics and Optimisation* (eds. J. Gabszewicz, J.F. Richard and L.A. Wolsey). Amsterdam: North-Holland.
- [8] Giudici, P., Green, P. 1999. Decomposable graphical Gaussian model determination. *Biometrika* 86, 785-801.
- [9] Gonzalez-Farias, G., Dominguez-Molina, J., Gupta, A. 2004. The closed skew-normal distribution. In: Genton, M., ed. *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. Boca Raton, FL: Chapman Hall/CRC, 25-42.

- [10] Højsgaard, S., Edwards, D., Lauritzen, S. 2012. Graphical models with R. Springer-Verlag, New York.
- [11] Khare, K., Rajaratnam, B. 2011. Wishart distributions for decomposable covariance graph models. *Annals of Statistics* 39, 514555.
- [12] Lauritzen, S.L. 1996. Graphical models. Oxford University Press, New York.
- [13] Letac, G., Massam, H. 2007. Wishart distributions for decomposable graphs. *Annals of Statistics* 35, 1278-1323.
- [14] Mardia, K.V., Kent, J.T., Bibby, J.M. 1979. Multivariate analysis, New York: Academic Press.
- [15] Mouchart, M. 1976. A note on Bayes theorem. *Statistica* 36, 349-357.
- [16] Newton, M.A., Raftery, A.E. 1994. Approximate Bayesian inference by the weighted likelihood bootstrap(with discussion). *Journal of the Royal Statistical Association* 3, 3-48.
- [17] Paulsen, V.I., Power, S.C., Smith, R.R. 1989. Schur products and matrix completions. *Journal of Functional Analysis* 85, 151-178.
- [18] Rue, H., Held, L. 2005. Gaussian Markov random fields: Theory and Applications. London: Chapman and HallCRC Press.
- [19] Whittaker, J. 1990. Graphical models in applied multivariate statistics. Wiley, Chichester.